

# DATA LITERACY ESSENTIALS

FSS 2022

## Data Literacy Essentials: Daten verstehen

Ulrich Krieger

Vielen Dank Dr. Ira Maschmeier für die Vorarbeit an diesem Kurs

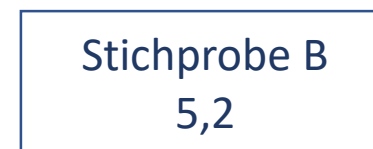
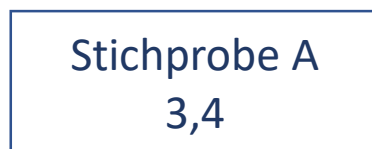
# 1. Beispiel: die Bewertung der CDU

## Vorfrage Gruppe A

„Wissen Sie zufällig, welcher Partei Hans-Georg Maaßen angehört?“

2 Stichproben werden befragt, jeweils 100 Personen:

„Alles in allem, was halten Sie ganz allgemein von der CDU?“



(Adaptiert nach Schwarz und Bless 1992; Porst 2000)

## 2. Beispiel: Erfolg im Leben

2 Stichproben werden befragt, jeweils ca. 500 Personen:  
**„Wie erfolgreich waren Sie in Ihrem Leben?“**

Auf einer Skala von

### Version A

0 = überhaupt nicht erfolgreich bis  
 10 = außerordentlich erfolgreich

### Version B

-5 = überhaupt nicht erfolgreich  
 5 = außerordentlich erfolgreich

<p>Version A 6,4</p>
--------------------------

<p>Version B 7,3</p>
--------------------------

(Adaptiert nach Petersen, 2002)

“We can be blind to the obvious,  
and we can be blind to our blindness”

Daniel Kahnemann

# Empirische Wissenschaftsdisziplinen



# Daten verstehen im Forschungsprozess



# Arten von Forschungsfragen



HWS 2021/22

Die Art der Forschungsfrage entscheidet über die Interpretierbarkeit und Aussagekraft der Ergebnisse.

## Thema: Ernährung und Viruserkrankungen

- **Deskriptiv:** Beschreibend
- **Exploratorisch:** Hypothesen-generierend
- **Inferentiell:** Aussagen über Grundgesamtheit
- **Vorhersagende:** Prognostizierend
- **Kausal:** einen Zusammenhang herstellend

# Eine „gute“ Forschungsfrage...

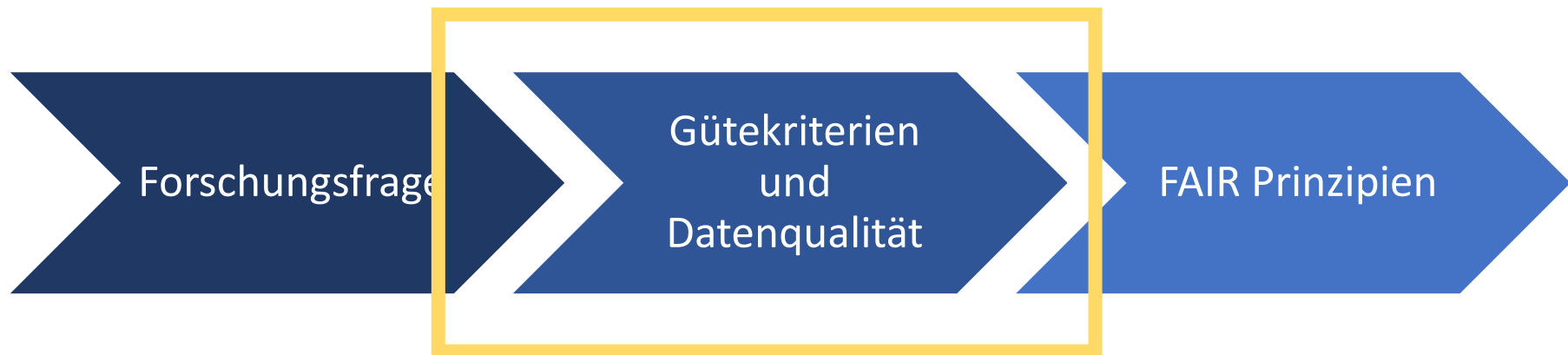


HWS 2021/22

1. ... ist von Interesse für mein Publikum
2. ... wurde noch nicht erforscht
3. ... beruht auf einem plausiblen theoretischen Rahmen
4. ... sollte beantwortbar sein
5. ... ist spezifisch



# Daten verstehen im Forschungsprozess



## Objektivität

- Unabhängigkeit von Befunden von der Untersuchungssituation
- Durchführung & Auswertung personenunabhängig : Intersubjektive Nachvollziehbarkeit
- Lösung: beispielsweise standardisierte Manuale

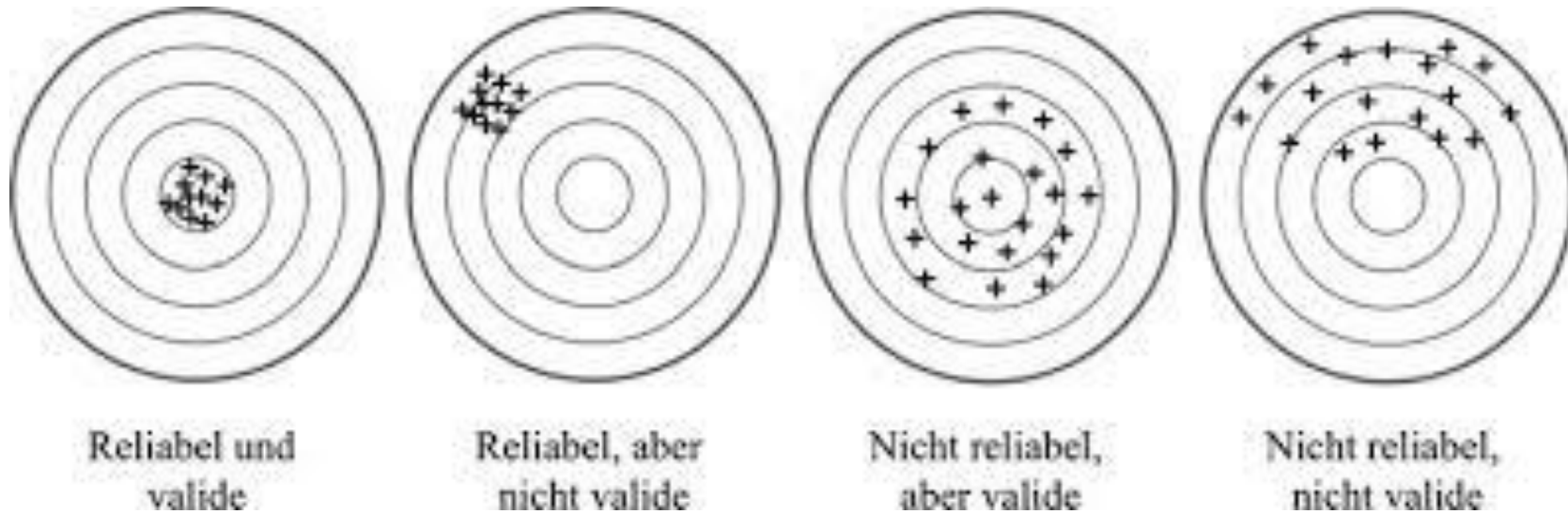
## Reliabilität

- Reproduzierbarkeit der Messergebnisse = dasselbe Ergebnis bei Messwiederholung
- Lösungsansätze: Replikation, Paralleltest, Split-Half

## Validität

- Gültigkeit = Grad der Genauigkeit, mit der ein Test das misst, was er messen soll
- Überprüfungsansätze:
  - Inhaltsvalidität
  - Kriteriumsvalidität
  - Konstruktvalidität
- Erprobte Skalen / Fragen nutzen

# Gütekriterien



(Heiss & Mattes, 2018)

# Datenaufbereitung und Fehlersuche



HWS 2021/22

[#DataQuality](#) is about cleaning up bad data – that has missing values, is incorrect or invalid in some way. But in order to **ensure data are trustworthy**, it is important to understand the key dimensions of data quality to assess how the data are “bad” in the first place.”

# Datenplausibilität



HWS 2021/22

ID	Name	Gebdat	Alter	Geschlecht	Telefon	PLZ	Email
56	Schmitt, Tina	28.07.1981	35	W	999-999	65197	Null
56	Dorothee Meschner	04.07.82	39	F	0211-473562	36999	Null
57	Hans Herrmann	10.03.78	43	M	089-67873	80331	hans@web.de

PLZ	Ort
65197	Wiesbaden
36996	Italien
80331	München

# Datenplausibilität

ID	Name	Gebdat	Alter	Geschlecht	Telefon	PLZ	Email
56	Schmitt, Tina	28.07.1981	35	W	999-999	65197	Null
56	Dorothee Meschner	04.07.82	39	F	0211-	36999	Null
57	Hans Herrmann	10.03.78	43				

PLZ	Ort
65197	Wiesbaden
36996	Italien
80331	München

## Achtsamkeit in Bezug auf

- Unterschiedliche Repräsentationen
- Widersprüchliche Werten
- Referenzen zwischen Datenquellen
- Vollständigkeit: fehlende Werte, Defaultwerte
- Duplikate
- Falsche oder unzulässige Werten
- Schreib- oder Tippfehler

# Datenqualität: Datenerhebung



HWS 2021/22

- Sponsor der Befragung
- Stichprobe, wer ist in der Studie?
  - Tests an Studierenden (WEIRD - Western, educated, industrialized, rich and democratic)
- Selbstselektion
  - Befragte selektieren sich selbst in die Befragung: Wer Lust hat, macht mit
  - Siehe z.B. <https://www.tagesschau.de/faktenfinder/umfrage-zweifelhaft-101.html>
- Repräsentativität: Eine inhaltsleerer Begriff
  - Statistisch nicht definiert
  - Basierend auf einer Zufallsstichprobe, aber viele Fehlerquellen
- Befragungen im Internet: Wer ist online?

Every Idiot can do an online survey ...

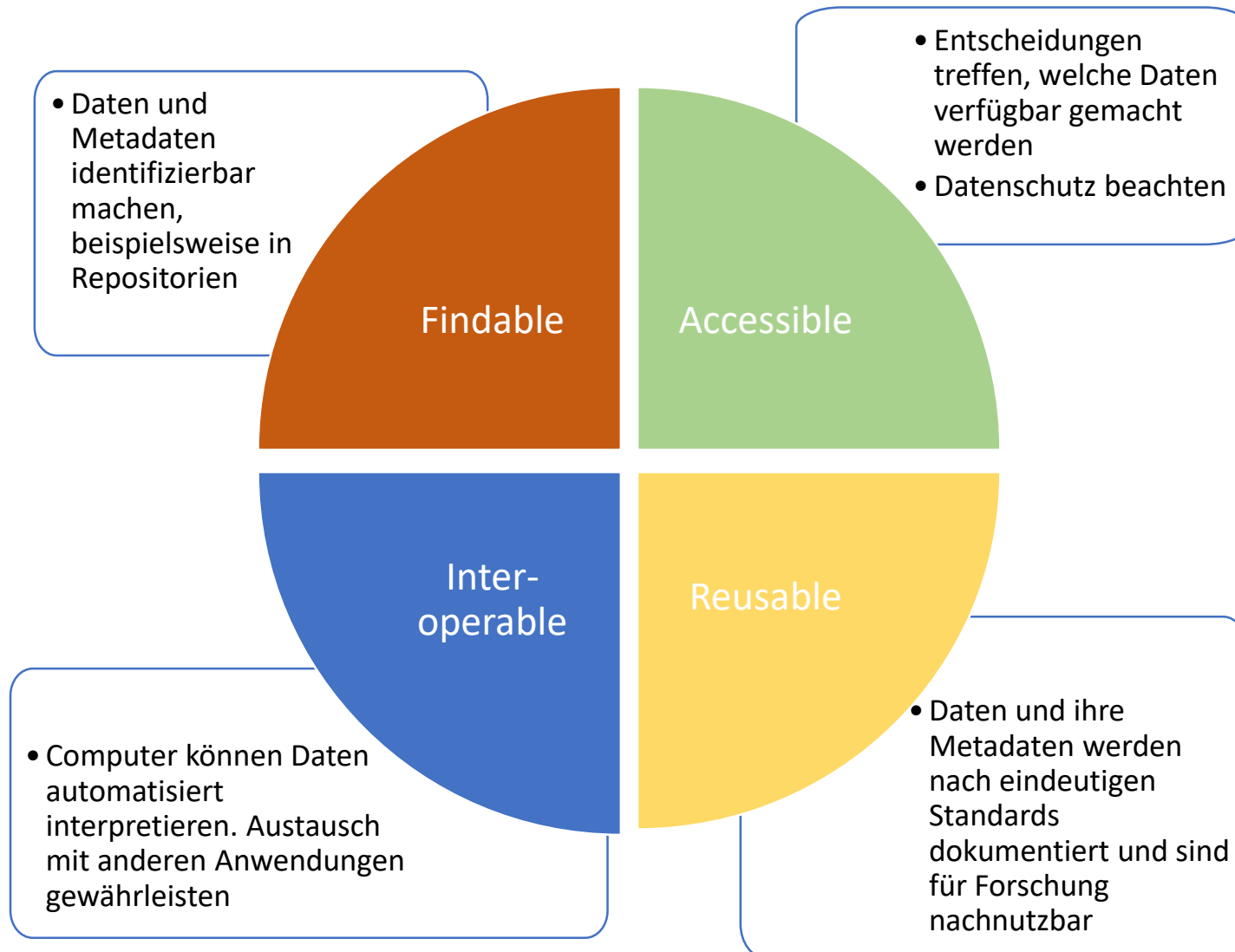
... and all the idiots are out there doing it!

# Daten verstehen im Forschungsprozess





# FAIR Prinzipien



# Daten verstehen



HWS 2021/22

und einordnen zu können, in der eigenen Forschung sowie bei der Bewertung fremder Forschungsergebnisse erfordert Wissen und Reflektionsvermögen.

# Literatur



HWS 2021/22

- Häder, M. (2019). *Empirische Sozialforschung*. Berlin, Germany: VS Verlag für Sozialwissenschaften. <https://doi.org/10.1007/978-3-658-26986-9>
- Peng, R. D., & Matsui, E. (2015). *The art of data science. A Guide for Anyone Who Works with Data*. Skybrude Consulting, LLC.

This Data Literacy Essential was developed as part of  
BERD – Business, Economic and Related Data.

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 460037581